ethnomusicology. However, trying to model the past audiences is akin to the performer/musicologist attempting to recreate the music of the past. Taruskin has summarised this best: '…what we call historical performance is the sound of now, not then. It derives its authenticity not from its historical verisimilitude, but from its being for better or worse a true mirror of late-twentieth century taste.'[5]

# Author's Response

## Modelling historical audiences: What can be inferred?

*Tuomas Eerola[1] and Marcus T. Pearce[2]*
[1]Durham University, Durham DH1 3RL
tuomas.eerola@durham.ac.uk
[2]Queen Mary University of London, London E1 4NS
marcus.pearce@qmul.ac.uk

*Introduction*

The target article ("Towards predictive models of music perception in historical audiences", TPM for short) made a bold interdisciplinary proposal and received a varied set of insightful commentaries from scholars in a range of disciplines, for which we are grateful. We have grouped the comments, and our responses, into three categories: first, those that relate to the cognitive modelling of music perception in general (i.e., regardless of time period); second, those that relate to difficulties in making inferences about historical listeners specifically; and third, those that relate to questions of contemporary musical structure and practice.

Before we proceed, we would like to make a few comments about interdisciplinary engagement. Mixing perspectives from separate disciplines such as historical musicology and systematic musicology is challenging due to differences in the underlying motivations, paradigms, terminology and nature of evidence commonly accepted in the disciplines. It is not so much a question of methods, techniques or analyses (these can be learnt and adapted) but whether dialogue between the different disciplines can be established and sustained (e.g., Volk & Honingh, 2012; Clarke, 2009) and whether the questions posed are meaningful across discplinary divides. In TPM, we approached questions that might usually be thought to fall under historical musicology with ideas from cognitive science, psychology and computational modelling. Our original work was not fully interdisciplinary as such, since TPM

[5] Richard Taruskin, "The Modern Sound of Early Music," in *Text and Act: Essays on Music and Performance* (NY - Oxford: Oxford University Press, 1995), 166. These ideas are also articulated in Richard Taruskin, "The Pastness of the Present and the Presence of the Past," in *Text and Act: Essays on Music and Performance* (NY - Oxford: Oxford University Press, 1995), 90-154.

represents mostly the view from systematic musicology, but the open peer commentaries have opened up a constructive discourse concerning the intersection of these disciplines. We hope that our responses below continue this spirit of interdisciplinary engagement.

*Modelling musical minds*

Our central claim from reviewing the existing literature on music perception is that musical expectations reflect a process of implicit learning through which (present-day) listeners acquire cognitive representations of structural relations in the music to which they listen. A second observation from the literature is that computational methods have been developed that allow such structural relations to be learned in an unsupervised manner and these methods have proved quite successful in simulating the expectations of present-day listeners. Our proposal in TPM is that it should be possible to simulate the cognitive representations of historical listeners using the very same computational mechanisms. The proposal rests on two assumptions. The first, is that we have enough relevant information about the music to which historical listeners were exposed. We consider this assumption in detail in a separate section below. The second assumption is that the cognitive process of implicit learning has remained intact over the time period in question (hundreds of years). Huron addresses this assumption, arguing that it would be surprising if earlier generations did not possess the ability, since it appears to be ubiquitous across species. Finally, it is important to note that within the proposed framework, implicit learning of musical structure could, in principle, be simulated by any unsupervised machine learning methods. We have focused on n-gram modelling because it has proved a powerful framework for modelling music perception in present-day listeners. However, future research will no doubt develop more refined models. Whether they can illuminate the musical perception of historical audiences, performers or composers, remains to be determined but many of the considerations discussed here will still apply.

D. Collins and Sequera raise the concern that research on musical expectation that we use to support our approach has been conducted on present-day listeners. To this point, Huron offers the insight that trying to understand the mind of a listener from the past is ontologically no different to trying to understand the mind of a present-day listener, regardless of their physical and cultural proximity. In both cases, understanding subjective experience is a practical process of inference based on the evidence available. The important difference concerns the nature of the evidence: with present-day listeners, we can test hypothesised cognitive mechanisms with empirical studies. For historical listeners, this is not possible but our framework is designed to attempt to make inferences from the data that is available, which is analogous to the way that contemporary music (e.g. pop music) has been used as data for accounting for expectations and tonality in present-day musical listeners (e.g., Temperley & Clercq, 2013).

Might it be possible to conduct empirical inferential experiments that bear on the question of historical music perception? Raman & Dowling suggest comparing the results of model simulations with responses of present-day musicians immersed in a

particular historical genre (e.g., Carnatic music). This is a highly useful suggestion although it is subject to assumptions, first, that the musical tradition itself has not developed over time and, second, that the perception of such listeners is immune from the experience of listening to music that was not available to historical listeners, and third that the emphasis of the culture-specific features is generally constant across cultures, which may not be the case (e.g., Fritz, 2013). Nonetheless, this is an interesting approach that is likely to produce convergent evidence. Another possibility for empirical testing would be to use a cognitive simulation to make predictions about historical transcription errors made by scribes, given evidence that present-day musicians tend to make transcription errors at points where expectations are disconfirmed (Unyk & Carlsen, 1987).

Huron points to evidence that the structural patterns that listeners represent are imperfect approximations of patterns that actually appear in the music they listen to. In other words, implicit statistical learning is subject to representational constraints on learning. More generally there might be other aspects of music perception that do not depend on implicit learning. If they are not dependent on musical experience, and we can assume that they have not changed in response to evolutionary pressures over the time period in question, then these aspects should be comparable between present-day and historical listeners. Therefore, it should be possible to incorporate them into the modelling framework by obeying such fundamental cognitive processing limitations as short-term memory, octave equivalence, categorisation of frequency onto hierarchically organised pitch levels and principles of auditory segregation (Stevens, 2004). Huron notes that music perception (historical or otherwise) might also reflect factors other than expectation, such as general ethological principles (presumably stable over time) and unique historically situated gestures that do not generalise (i.e., they are specific to a given period, culture and setting).

Conversely, there might be aspects of music perception other than expectation that are shaped by musical experience. If so, it should be possible to simulate these aspects using broadly the same framework we have outlined (though perhaps with differently parameterised models). In this respect, Harrison suggests looking at representations of musical timbre. This is an interesting suggestion although timbre perception is less well understood as a psychological phenomenon than pitch and the relevant aspects of instrumental and performance style may not be extant. However, given relevant historical evidence, this topic could well follow the same line of reasoning outlined in TPM where existing notions of how performers implicitly learn timbral relationships (cf. Tillmann & McAdams, 2004) could be applied to historical listeners. D. Collins suggests looking at emotional experience. Again this is an exciting possibility but emotional experience is an even more thorny area of music perception than expectation, so this might open up more questions and problems of interpretation than it solves. On the one hand, expectation is thought to play a role in the emotional experience of music but emotional experience is also thought to reflect individual episodic memories and learned associations with particular musical styles, amongst other mechanisms (see Juslin & Västfjäll, 2008), which are no longer extant for historical listeners. Could they be inferred?

Several commentators make useful observations regarding the *n*-gram modelling framework that we propose. Carter points to the limitations of fixed-order *n*-grams in terms of structural representation. This is quite correct though we note that these limitations can be addressed with variable-order Markov models, capable of combining information from models of different order and sophisticated multiple-viewpoint representational frameworks that allow models to combine information from multiple different musical features at different levels of representational abstraction (e.g. Conklin & Witten, 1995; Pearce, 2005). Raman & Dowling make an excellent suggestion in this respect to focus on contour and scale degree representations, which have been found to play an important role in memory for melody.

Lewis also proposes extending the approach with short-term modelling to simulate perception of repeated motives within musical works (e.g. Conklin & Witten, 1995; Conklin & Anagnostopoulou, 2006; Pearce, 2005). Although the proposed framework accommodates implicit learning of such intra-opus patterns using the short-term model, our initial simulations were limited to the long-term model for elegance of exposition. Future research should investigate this question directly.

Lewis also makes the useful point that while increasing the complexity of the model might offer advantages in terms of capturing musical structure, it also becomes more challenging to argue that it is a perceptual model. Therefore, research simulating historical listeners using the framework outlined in TPM must proceed in tandem with testing model developments against the musical perception of present-day musical listeners. It was partly for this reason that our simulations did not make use of feature combinations (as noted by T. Collins) even though this is possible within the modelling framework (by virtue of the multiple viewpoints representation scheme). Research has simulated present-day perception of music using multiple-viewpoint models (e.g. Hansen & Pearce, 2014) but doing so adds complexity. We thought that our illustrative examples would have greater clarity using single features. Nonetheless future research using the proposed framework should certainly exploit the full power of the multiple viewpoint representation scheme.

*Musical past*

A remark frequently made by the commentators (Huron, Lewis, Sequera, and T. Collins) is that we do not actually know with any degree of certainty what music historical audiences were exposed to. This is a fundamental issue to be addressed. However, it is not particular to the approach proposed in TPM, since it applies to any study of historical musical listeners. The music that has survived in scores and manuscripts probably represents only the tip of the iceberg constituting the music prevalent in each era, and quite likely the musical content remaining in manuscripts and collections is subject to biases of various kinds. This might be considered, on the one hand, an impasse or, alternatively, an interesting stimulus for pragmatic research to assess how much can be inferred with the available evidence. Take, for example, the question of how large a corpus one needs to simulate a listener from a given culture (historical or otherwise). This could be assessed empirically by using models

trained on different-sized corpora to simulate contemporary listeners from a given culture on a range of tasks, including tests of melodic expectation. However, this would not address difficulties relating to non-representative corpora resulting, for example, from the fact that much of the music that was heard in many historical periods has simply not survived. However, historical musicologists may be able to provide indirect indices of such music repertoire (using, for example, records from publishing, sales, concerts, marketplaces, private collections) that could inform model training by weighting the materials according to their assumed prevalence (see e.g., London, 2013) instead of relying on single instances within an existing corpora, as was done in the current implementation. Harrison recommends that the emphasis of the analysis should be placed after 1800 due precisely to the increased presence of such documentary evidence after this date. This would also present a fruitful opportunity for collaboration between historical musicologists and music psychologists although the inferential problem remains that there is still no full account (let alone recordings) of the music heard by 19th Century listeners.

Even if we could understand the ways that past listeners might have perceived music using well-studied cognitive processes (such as implicit learning, expectation and auditory stream segregation), the evidence in terms of musical materials used to train and evaluate the simulations is subject to various interpretations (leaving open the possibility of misinterpretation). Ceulemans, for example, questions the relevance of using key profiles for analysing music that is fundamentally modal. In TPM, we followed a data-driven approach reflecting the way the pitches (and intervals) within the octave are used in any given corpus, rather than imposing a Major or Minor tonal hierarchy as a universal solution. This rests on the assumption that most scale systems contain hierarchies of tones, reflected in the statistical structure of the music, which should be learnable, regardless of whether the music is modal or tonal (Huron & Veltman, 2006). Future research should investigate this assumption by comparison with explicit representations of tonal and modal pitch representations, both of which can be accommodated with multiple viewpoints. Ceulemans also provides an interesting proposal for how TPM might be applied to resolve the question of why the *final* and *repercussa* are not always a fifth apart, or how such ambiguous endings emerge across history in conjunction with the more functional role of harmony.

In a similar way, D. Collins raises doubts about the usefulness of the analysis applications (authorship, style classification, etc.) due to many potential pitfalls in the process. Naturally pitfalls exists but we do think that many interesting research questions will emerge from a consideration of how best to navigate them and that fruitful answers to those questions will only result from interdisciplinary dialogue – a good example is the suggestion by D. Collins of using text setting to identify modes of listening and avoid inappropriate generalisation across those modes.

*Musical context*

Finally, we address issues of contemporary musical practice or stylistic sensitivity. Harrison points out correctly that the representations used in the simulations must be sensitive to cross-cultural differences and Carter notes, specifically, that Implication-

Realisation models (Narmour, 1990; 1992) only apply to melodic expectations in Western tonal styles. The IR model consists of two systems – a bottom-up system consisting of a set of universal rules of melodic implication and a top-down system, which is sensitive to experience and, therefore, potentially variable between musical styles (extra-opus effects) and pieces (intra-opus effects). In fact, the bottom-up principles tend to reflect regularities in actual music (Thompson & Stainton, 1996) and some, such as pitch proximity are apparent both in the music and in the expectations of listeners in non-Western cultures (Carlsen, 1981; Eerola, Louhivuori, & Lebaka, 2009; Huron, 2001; Krumhansl et al., 2000). Therfore, they may in fact reflect universal physical constraints of performance, such as the difficulty of performing large intervals or tessitura constraints (Russo & Cuddy, 1999), which are subsequently acquired by listeners via implicit statistical learning through exposure. Our approach does not use the IR principles but rather takes advantage of such a process of implicit statistical learning through exposure, without making a distinction between top-down or bottom-up effects (everything is, in effect, top-down and dependent on experience). This means that our approach is capable, in principle, of simulating the expectations of listeners from other cultures or points in history as a function of the music to which they were exposed.

T. Collins highlights the fact that vertical constraints exist between voices and Lewis notes, more specifically, that multipart writing places constraints on the inner voices. It seems likely that such vertical constraints are represented and processed in musical listening, though further research with present-day listeners is required to develop an understanding of exactly how. For these reasons, though understandable given the present state of knowledge, our treatment of each voice independently is inadequate as a representation of musical structure. We acknowledge this and see it as a spur to further research on representation of polyphonic structure in music that is amenable to modelling using unsupervised learning methods of the kind we described. As noted in TPM, we believe that cognitive models of stream segregation (Bregman, 1990) are likely to be useful in identifying the parallel streams of notes and chords that listeners identify in listening to polyphonic music.

*Conclusion*

The target article, the commentaries and this response represent a dialogue between several disciplines of music research. We have attempted to outline an approach to understanding historical musical listeners using empirical tools and methods from the sciences. This is very much the beginning of an interdisciplinary research programme and subsequent developments will establish how successful the approach proves to be. It is certain, however, that if it is to be successful then it will require the collaboration of experts in computational musicology, systematic musicology, historical musicology amongst other disciplines. Therefore, we think this an appropriate point to reflect on the nature of the interdisciplinary dialogue that is likely to prove fruitful.

One thing required is a patient dedication to the task, an acceptance that it takes time to break-down disciplinary boundaries so as to allow true collaboration and a

realisation that different motivations and methods can co-exist side by side and even complement each other. This article itself has been the subject of many discussions over a period of about five years. The research started in 2011 with invitations by Richard Parncutt to contribute to an exploratory workshop on "Cognition of Early Polyphony" which took place in March 2012, funded by the European Science Foundation (ESF). We prepared our own separate case studies of how statistical models of music could be applied to Renaissance music and negotiated the challenge of presenting them to other scholars from such fields as historical musicology, ethnomusicology as well as music cognition and neuroscience. The workshop itself represented an exciting mixture of rich, interdisciplinary engagement and challenging interpretation of unfamiliar goals, terminology and methodological approaches.

In our case, this could well have been the end of the story and a return to our own disciplinary pursuits. However, thanks to the encouragement and support of the editorial team (FW, BT), we decided to venture into an attempt to deliver a new perspective on understanding the perception of historical listeners. In fact, TPM already represents a dialogue between our respective approaches since, at the outset, one of us had focussed on theoretical modelling frameworks, and the other on the application of models to Renaissance music. The review process soon brought back the broader frame of reference already at play in the workshop. The shortcomings of TPM were neatly laid out by the reviewers in three primary areas: (a) the sophistication of musical processing; (b) treatment of aspects of music history; and (c) consideration of the musical context. Similar issues have been raised and elaborated upon in the open peer commentaries and we are grateful for the opportunity to respond to these issues, thereby furthering the interdisciplinary dialogue. In our opinion, this iterative process has proved highly insightful and, on many occasions, served as a reminder of how to engage and communicate with music researchers outside our specific discipline. On a personal note, we often find ourselves involved with other scientific disciplines (e.g., psychology, cognitive science, computer science, neuroscience) which share many methods, concepts and terminology. To be involved in a constructive discussion of research questions between the humanities and the sciences has proved overwhelmingly more challenging but also far more rewarding in terms of the knowledge and understanding that can result.

## References

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, *1*(1), 12–29.

Clarke, E. (2009). Boundaries, expectations and empirical research: A commentary on Judith Becker's "Crossing Boundaries". *Empirical Musicology Review*, *4*(2), 71–74.

Conklin, D., and Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *INFORMS Journal on Computing*, *18*(3), 285–293.

Conklin, D., and Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*, 341-354. doi:10.1037/0033-295X.85.4.341

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception, 3*, 281-296. doi:10.2307/40285338

Dowling, W. J., and Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*, 30-49. doi:10.1037/h0094275

Dowling, W. J., Kwak, S.-Y., and Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57*, 136-149. doi:10.3758/BF03206500

Dowling, W. J., Raman, R., Ramesh, A., and Tillmann, B. (2015, November). *The time course of responses to modulations in classical music*. Poster session presented at the meeting of Psychonomic Society, Chicago, IL.

Dowling, W. J., & Tillmann, B. (2014). Memory improvement while hearing music: Effects of structural continuity on feature binding. *Music Perception, 32*, 11-32. doi:10.1525/mp.2014.32.1.11

Eerola, T., Louhivuori, J., and Lebaka, E. (2009). Expectancy in North Sami yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae*, *13*(2), 39–70.

Fritz, T. (2013). The dock-in model of music culture and cross-cultural perception. *Music Perception: An Interdisciplinary Journal*, *30*(5), 511–516.

Gooley, D. (2004). *The Virtuoso Liszt*. Cambridge: Cambridge University Press.

Gur, G. (2011). The Spectre of the End: Musical Avant-Gardism and the Philosophical Turn. *International Review of the Aesthetics and Sociology of Music*, *42*(2), 267-284.

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*(1), 1–64.

Huron, D., and Veltman, J. (2006). A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, *1*, 170–177.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.

Knopke, I. and Jürgensen, F. (2011). Symbolic data mining in musicology. In T. Li, M. Ogihara, and G. Tzanetakis (Eds.), Music Data Mining (pp. 327-345). San Diego, CA: CRC Press.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*(1), 13–58.

London, J. (2013). Building a representative corpus of classical music. *Music Perception*, *31*, 68–90.

Mangani, M. and Sabaino, D. 2008. "Tonal types and modal attributions in late Renaissance polyphony: new observations", *Acta Musicologica* 80/2

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realisation model*. Chicago: University of Chicago Press.

Nattiez, J. (1990). *Music and Discourse: Toward a Semiology of Music*. Translated by Carolyn Abbate. Princeton: Princeton University Press.

Pearce, M. T., and Eerola, T. (in press). Towards predictive models of music perception in historical audiences. *Journal of Interdisciplinary Music Studies*.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (PhD thesis). Department of Computing, City University, London, UK.

Raman, R., and Dowling, W. J. (2016). Real-time probing of modulations in South Indian classical (Carnātic) music by Indian and Western musicians . *Music Perception, 33,* 367-393.

Russo, F. A., & Cuddy, L. L. (1999, March). A common origin for vocal accuracy and melodic expectancy: Vocal constraints. Berlin, Germany: Paper presented at the Joint Meeting of the Acoustical Society of America and the European Acoustics Association.

Stevens, C. (2004). Cross-cultural studies of musical pitch and time. *Acoustical Science and Technology*, *25*(6), 433–438.

Temperley, D., and Clercq, T. de. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, *42*(3), 187–204.

Thompson, W. F., & Stainton, M. (1996). Using *humdrum* to analyse melodic structure: An assessment of Narmour's implication-realisation model. *Computing in Musicology*, *12*, 24–33.

Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustic (dis)similarities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(5), 1131–1142.

Toiviainen, P., and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception, 32*, 741-766. doi:10.1068/p3312

Unyk, A. M., and Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, *7*(1), 3–23.

Volk, A., and Honingh, A. (2012). Mathematical and computational approaches to music: Challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music*, *6*(2), 73–81.

Volk, A., and de Haas, W. Bas. (2013). A corpus-based study on ragtime syncopation. In Proc. ISMIR (pp. 163-168). Curitiba, Brazil.

Wiering, F. (2001). *The Language of the Modes: Studies in the History of Polyphonic Music*, New York: Routledge.